

DISK CACHE CONTROL SYSTEM

Patent Number: JP5189314
Publication date: 1993-07-30
Inventor(s): TOURAKU MAMORU; others: 04
Applicant(s): HITACHI LTD
Requested Patent: ☐ JP5189314
Application Number: JP19920001279 19920108
Priority Number(s):
IPC Classification: G06F12/08; G06F3/06; G06F12/08; G11B20/10
EC Classification:
Equivalents:

Abstract

PURPOSE: To prevent the deterioration of a system through put by storing data in one of disk caches at the time of transmitting the data stored in a storage device to a channel.

CONSTITUTION: The disk cache is divided into a cache 1-329 and a cache 2-330, and at the time of reading the data from a magnetic disk device(DKU) 334, the data are stored in one cache 2-330. Thus, even at the time of the occurrence of a fault at the cache 1-329 the reading of the data can be executed from the cache 2-330 afterwards. And also, at the time of the occurrence of the fault at the cache 2-330, the data are reduced from the cache, the next reading is executed from the DKU 334, and the read data are transmitted to a channel device(CHA) 302 and the existing cache 1-329, and stored. At that time, almost the half(average value) of the entire data are stored in one of the caches 329 and 330, and the data transfer between the CHA and the cache can be attained.

Data supplied from the esp@cenet database - I2

HEI 5-189314

[DETAILED DESCRIPTION OF THE INVENTION]

[0001]

[Industrial Application Field] The present invention relates to, in a disk control system equipped with a disk cache memory, an arrangement of a disk cache and a method of storing data through the use of a plurality of disk caches.

[0002]

[Description of Related Art] The following technique has been known as a conventional technique. FIG. 2 is an illustration of an example of an arrangement of a conventional system.

[0003] The system is made up of a central processing unit 101, channel devices (each of which hereinafter will be referred to shortly as a "CHA") 102 to 109, a disk control unit (which hereinafter will be referred to shortly as a "DKC") 110, and a magnetic disk device (which hereinafter will be referred to shortly as a "DKU") 131.

[0004] In the DKC 110, microprocessors 1 (which hereinafter will be referred to shortly as "MP1") 112 to 119 are connected which conduct the processing on channel switches 111, 120 and the CHAs 102 to 109, and microprocessors 2 (which hereinafter will be referred to shortly as "MP2") for controlling the DKU 131, a disk cache (which hereinafter will be referred to shortly as a "cache") 129 a battery-backup non-volatile memory (which hereinafter will be referred to shortly as an "NVS") 128

are connected to the MP1. In the illustration, dotted lines 121 to 123 denote a power supply region.

[0005] The NVS 128 is for storing data at the occurrence of a writing instruction (which hereinafter will be referred to shortly as "WRprocessing") from a host device. The MP2 (124 to 127) writes the data of the NVS 128 in the DKU 131 at the occurrence of a trouble of the cache 129 or after the re-start of the system.

[0006] The data flow in this system is as follows.

[0007] Let it be assumed that a processing request from the central processing unit 101 is implemented via the CHA 102 and the channel switch 111 with respect to the MP1 (112).

[0008] At the occurrence of a readout instruction (which hereinafter will be referred to shortly as "RDprocessing") from the host device, the MP 1 (112) makes a decision as to whether or not the target data exists in the cache 129. If the data does not exist in the cache 129, the RDprocessing is conducted with respect to the DKU 131 and this data is simultaneously transferred to the CHA 102 and the cache 129.

[0009] In a case in which this data exists in the cache 129, the RD processing is not conducted with respect to the DKU 131, but the data in the cache 129 is transferred to the CHA 102. If the data exists in the cache 129, the time for positioning of the DKU 131 and the like become unnecessary, and because of no dependence on the performance of the DKU 131, the transfer speed becomes high, thus achieving the speed increase.

[0010] A concrete example is shown in FIG. 3. FIG. 3 is an illustration of a track format of the DKU 131. Index markers 201 and 202 exist on a track, and represent a start point and an end point, respectively. On the track, data are stored in the form of data A to D 203 to 206. In a case in which the data A 203 is read out through the RD processing (naturally, the RD processing has not been conducted previously with respect to this track), although the data A 203 is transferred to the CHA 102 and the cache 129 and the MP 1 (112) then informs the CHA 102 of the completion, the data B to D 204 to 206 (all the data before the index marker 202) after this data are sent to the cache 129 to be stored in the cache 129. This is a function to successively store data in the cache in range from the processed track to several tracks ahead so that the target data always exists in the cache at the RD processing on the subsequent tracks. Accordingly, in the RD processing on the data subsequent to the data A 203 after this operation, the data in the cache 129 is transferred to the CHA 102, thus improving the system throughput.

[0011] There are a method in which the data is directly written in the DKU 131 at the RD processing and a method in which the data is written in the cache 129 and the completion is notified to the CHA 102 at this time and, following this, the data is written in the DKU 131 in an asynchronous relation to the host device. In the case of the latter method, since the DKU 131 does not undergo the processing during the connection with the CHA 102, the time for the positioning and the like becomes unnecessary, thereby shortening the response time to the

host device accordingly. However, usually, the cache 129 is not backed up by the battery and, if a trouble occurs in the cache 129, the data written can disappear. The NVS 128 is provided in order to prevent this. The writing data from the host device is stored in the cache 129 and further in the NVS 128, thereby keeping the data even if a trouble occurs in the cache 129 or even at the power-off. At the power-on after the power supply disconnection, the data in the NVS 128 can be transferred to the DKU 131.

[0012] In the case of a system which conducts the WR processing frequently, it is considered that the NVS 128 falls into a full condition. At this time, it stands ready to the RD processing from the host device and waits until a free area appears in the NVS 128, and then resumes the operation. Therefore, the system performance lowers after the NVS 128 falls into a full state.

[0013] In a case in which a trouble occurs in the cache 129, under an environment of 24-hours operation, the cache 129 is closed and an object such as package exchange is conducted. The cache 129 cannot be used during this time, and all the processing from the CHA 102 are conducted directly with respect to the DKU 131. This lowers the processing speed. Moreover, also restoration of the cache 129, the performance extremely degrades until the data is accumulated in the cache 129.

[0014] The technique disclosed in Japanese Patent Laid-Open No. HEI 2-90313 is such that a selection standard for a storage path in a disk control unit is prepared and a storage path which is in a non-used condition (the processing is not conducted yet) is used preferentially

to increase the using frequency of a disk cache for realizing the processing speed increase. In the arrangement disclosed in this well-known example, in a case in which a trouble occurs in the disk cache, the fast data transfer using the disk cache becomes completely impossible.

[0015]

[Problems to be Resolved by the Invention] In the case of the above-mentioned conventional techniques, after the non-volatile memory and NVS, backed up by a battery, reach the full condition, the system performance degrades.

[0016] Moreover, during a time period from when a trouble occurs in the disk cache until the exchange and restoration of the cache memory which is in trouble, the effects of the disk cache disappear and the system throughput lowers extremely. In addition, also after the power-on is again made, since the accumulation of the data in the disk cache is made from the beginning, the system throughput drops during this time.

[0017] Therefore, the present invention has been developed in order to remove these problems, and the present invention provides, in a storage system including a disk cache mounted disk control unit, a disk cache control method suitable for preventing the drop of the system throughput.

[0018]

[Means for Resolving the Problems] A plurality of disk caches, divided, are provided in a disk cache mounted disk control unit. The data transfer paths for the plurality of disk caches are placed independently. Moreover, each

of the disk caches is battery-backed up, and independent power supply lines are placed with respect to these disk caches.

[0019] The readout data from a storage unit is stored in an arbitrary one of the disk caches and writing data in the storage unit is likewise stored in arbitrary two of the disk caches for duplexing.

b)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平5-189314

(43) 公開日 平成5年(1993)7月30日

(51) Int.Cl. ⁵	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 12/08	3 2 0	7232-5B		
3/06	3 0 4 E	7165-5B		
12/08	J	7232-5B		
G 1 1 B 20/10	D	7923-5D		

審査請求 未請求 請求項の数1(全 6 頁)

(21) 出願番号 特願平4-1279

(22) 出願日 平成4年(1992)1月8日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 東落 守

神奈川県小田原市国府津2880番地株式会社
日立製作所小田原工場内

(72) 発明者 倉野 昭

神奈川県小田原市国府津2880番地株式会社
日立製作所小田原工場内

(72) 発明者 竹内 久治

神奈川県小田原市国府津2880番地株式会社
日立製作所小田原工場内

(74) 代理人 弁理士 小川 勝男

最終頁に続く

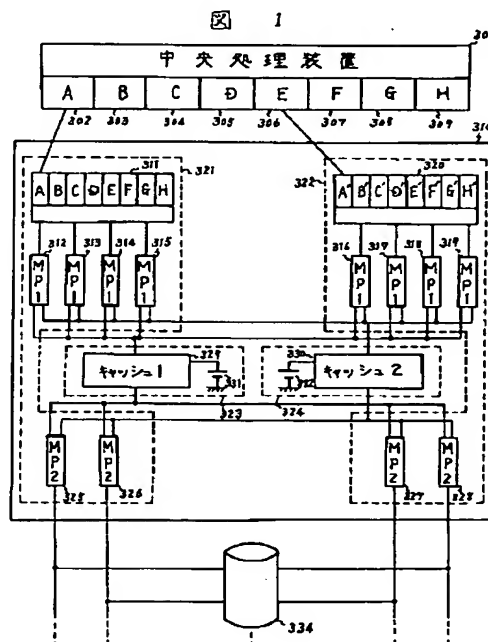
(54) 【発明の名称】 ディスクキャッシュ制御方式

(57) 【要約】

【目的】 ディスクキャッシュを搭載したディスク制御装置を含む記憶システム内において、システムスループットの低下を防ぐ。

【構成】 ディスク制御装置内に、バッテリーバックアップされたディスクキャッシュを複数設ける。磁気ディスク装置からの読み出しデータは、ディスクキャッシュの任意の一方に記憶させ、チャネルからの書き込みデータは、2つのディスクキャッシュに記憶させる。

【効果】 ディスクキャッシュの機能を損なうことなく、継続して使用することができる。



【特許請求の範囲】

【請求項1】 上位装置とデータの入出力を行うチャネルと、データを記録再生する記憶装置と、前記チャネルと前記記憶装置の間に設けられ前記記憶装置を制御する制御装置を含む計算機システムにおいて、前記制御装置内に複数のディスクキャッシュを設け、前記チャネルから送られたデータを前記記憶装置に書き込むときは、複数のディスクキャッシュにデータを二重に格納し、前記記憶装置に記憶されたデータを前記チャネルに送るときは、前記ディスクキャッシュの1つにデータを格納することを特徴とするディスクキャッシュ制御方式。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明はディスクキャッシュメモリを備えるディスク制御装置における、ディスクキャッシュの構成、及び複数のディスクキャッシュを用いたデータの記憶方式に関する。

【0002】

【従来の技術】 従来技術として次に示す技術がある。図2は例としてあげる従来システムの構成図である。

【0003】 システムは中央処理装置101、チャネル装置（以下「CHA」と略称する。）102～109、ディスク制御装置（以下「DKC」と略称する。）110、及び磁気ディスク装置（以下「DKU」と略称する。）131より構成される。

【0004】 DKC110は、チャネルスイッチ111、120とCHA102～109との処理を行うマイクロプロセッサ1（以下「MP1」と略称する。）112～119が接続されていて、DKU131の制御を行うマイクロプロセッサ2（以下「MP2」と略称する。）とディスクキャッシュ（以下「キャッシュ」と略称する。）129と、バッテリーバックアップされて不揮発なメモリ（以下「NVS」と略称する。）128がMP1と接続され構成される。図中点線121～123は電源領域を表す。

【0005】 NVS128は上位装置からの書き込み命令（以後「WR処理」と略称する。）時のデータを記憶するものである。MP2（124～127）はキャッシュ129障害時、又はシステム再始動後、NVS128のデータをDKU131に書き込む。

【0006】 本システムのデータの流れは以下のとおりである。

【0007】 中央処理装置101からの処理要求は、CHA102、チャネルスイッチ111を経由してMP1（112）に対し行なわれるとする。

【0008】 上位装置からの読み出し命令時（以後「RD処理」と略称する。）には、MP1（112）はキャッシュ129内に目的のデータが存在するか否かを判定する。キャッシュ129内にデータが存在しない場合、D

KU131へ対しRD処理を行い、当該データをCHA102とキャッシュ129へ同時に転送する。

【0009】 当該データがキャッシュ129内に存在する場合、DKU131へのRD処理は行わず、キャッシュ129内のデータをCHA102へ転送する。キャッシュ129内にデータが存在すれば、DKU131の位置決め時間等が不要となり、又DKU131の性能に依存しないため転送速度も高速となり、処理の高速化が図れる。

10 【0010】 具体例を図3を用いて示す。図3はDKU131のトラックフォーマットである。トラック上にはインデックスマーカ201、202があり、トラックの始点、終点を表す。トラック上にデータはデータA～D203～206という形で記憶されている。RD処理によりデータA203を読み出す場合（当然、以前に当該トラックへのRD処理を行っていない。）データA203をCHA102、キャッシュ129へ転送し、その後MP1（112）はCHA102に対し終了報告を行うが、キャッシュ129には当該データ以降のデータB～D204～206（インデックスマーカ202迄のデータ全て）を送り、キャッシュ129内に記憶させる。これは、処理を行ったトラックから数トラック先まで連続してキャッシュ内にデータを記憶させ、以降のトラックのRD処理時、目的データが常にキャッシュ内に存在する様にする機能である。これにより、本動作以降のデータA203以降のデータのRD処理では、キャッシュ129内のデータをCHA102に転送することとなりシステムスループットを向上させることができる。

30 【0011】 WR処理時、直接DKU131にデータを書き込む方法と、キャッシュ129内にデータを書き込み、この時点でCHA102に終了報告を行いその後上位装置とは非同期にDKU131へデータを書き込む方法がある。後者の方が、CHA102との接続中にDKU131を処理しない為、位置付け等の時間が不要な分、上位装置に対する応答時間を短縮することができる。しかし、通常キャッシュ129はバッテリーバックアップされておらず、キャッシュ129に障害が発生すると書き込んだデータが消滅してしまう。これを防止する為にNVS128を用意している。上位装置からの書き込みデータをキャッシュ129と共にNVS128に記憶させる。これによりキャッシュ129に障害が発生しても、又電源OFFとなってもデータは保障される。電源断後の電源投入後にNVS128内のデータをDKU131に反映させることができる。

【0012】 WR処理の多いシステムではNVS128が満杯となる場合も予想される。この時、上位装置からのWR処理を一時待機させ、NVS128に空きエリアができるまで待ち、その後再開させる。この為、NVS128が満杯となった後はシステムの性能が低下する。

50 【0013】 キャッシュ129に障害が発生した場合、

24時間運転の環境ではキャッシュ129を閉塞し、パッケージ交換等の対象を行う。この間キャッシュ129は使用できず、CHA102からの処理は全てDKU131に対して直接行うこととなる。これでは処理速度が低下する。又、キャッシュ129が回復した後もキャッシュ129内にデータが蓄積されるまでの間、性能が著しく低下する。

【0014】特開平2-90313号に記載の技術では、ディスク制御装置内のストレージバスの選択基準を設け、未使用状態の(処理を実行していない)ストレージバスを優先して使用することにより、ディスクキャッシュの使用頻度を上げ、処理の高速化を実現している。本公知例の構成では、ディスクキャッシュに障害が発生した場合、ディスクキャッシュを用いた高速データ転送が完全に不可能となる。

【0015】

【発明が解決しようとする課題】前記従来技術では、バッテリーバックアップされて不揮発なメモリ、NVSが満杯となった後はシステムの性能が低下する。

【0016】また、ディスクキャッシュに障害が発生してから障害キャッシュメモリの交換回復までの間、ディスクキャッシュの効果が失われシステムスループットが著しく低下する。そして電源再投入後も、ディスクキャッシュへのデータの蓄積を最初から行う為、この間もシステムスループットが低下する。

【0017】そこで本発明はこれらの問題を解決するためになされたもので、ディスクキャッシュを搭載したディスク制御装置を含む記憶システム内において、システムスループットの低下を防ぐために好適なディスクキャッシュの制御方式を提供するものである。

【0018】

【課題を解決するための手段】ディスクキャッシュを搭載したディスク制御装置内に複数に分割したディスクキャッシュを設ける。この複数のディスクキャッシュに対するデータ転送用のバスも独立とする。また、それぞれのディスクキャッシュをバッテリーバックアップし、これらに対し独立した電源ラインを置く。

【0019】記憶装置からの読み出しデータは、複数のディスクキャッシュの任意の1つに記憶させ、記憶装置への書き込みデータは同様に任意の2つのディスクキャッシュに記憶させ2重化する。

【0020】

【作用】バッテリーバックアップにより不慮の電源OFF後も直前のディスクキャッシュ状態が保持され、電源再投入直後からディスクキャッシュを用いた処理が可能となる。又、書き込みデータはディスクキャッシュ内に保持されている為、電源投入後、記憶装置に反映させることができる。

【0021】この様にディスクキャッシュを実質的に複数に、独立的に分割することにより、ディスクキャッシ

ユの1つに障害が発生してもディスクキャッシュ効果を失うことなく保守を行うことが可能となる。又、書き込みデータは2重化する為、消滅することはない。

【0022】

【実施例】以下、本発明の一実施例を図を用いて説明する。

【0023】図1は本発明における実施例のシステム構成図である。本実施例ではディスクキャッシュを2分割とする。(キャッシュ1(329)、キャッシュ2(330))MP1(312~319)、MP2(325~328)とキャッシュ1(329)、キャッシュ2(330)間はキャッシュ単位にバスを設ける。又、キャッシュ1(329)、キャッシュ2(330)にバッテリー331、332を設けバッテリーバックアップとしている。電源領域321~324単位に電源を供給する。

【0024】処理は従来技術と同様に中央処理装置301、CHA302、チャネルスイッチ311、MP1(312)により処理を行う。DKU334の制御はMP2(325)で行う。

【0025】上位装置からのRD命令時、MP1(312)はキャッシュ(329、330)内に目的データが存在するか否か確認する。事前に当該データ(トラック)が読み込み(又はWR処理)済みでキャッシュ内(ここではキャッシュ2(330)とする。)にデータが存在すれば、キャッシュ2(330)よりCHA302へデータ転送を行う。キャッシュ1(329)、キャッシュ2(330)内に目的データが存在しない場合、DKU334よりデータを読み出しCHA302へ転送する。DKU334よりデータを読み出す時、従来の装置と同様にキャッシュにデータを記憶させるが、この時、一方のキャッシュにのみデータを記憶させる。これはMP2(325)で選択する。本例ではキャッシュ2(330)にデータを記憶したとする。

【0026】もし、キャッシュ1(329)に障害が発生してもデータはキャッシュ2(330)に記憶されている為、以降当該データの読み出しはキャッシュ2(330)より行うことができる。

【0027】キャッシュ2(330)に障害が発生した場合、当該データはキャッシュより消滅してしまう為、次の読み出しはDKU334より行う。読み出したデータはCHA302と生存しているキャッシュ1(329)に送られ記憶される。

【0028】2分割されたキャッシュ(329、330)の一方には、キャッシュに格納すべきデータ全体の約1/2(平均値)が格納されており、一方のキャッシュで障害が発生しても他方のキャッシュで記憶されているデータの処理は、CHとキャッシュとのデータ転送が可能であり処理速度は低下しない。よって一方のキャッシュに障害が発生してもキャッシュ効果はゼロとならない。(約1/2となる。)障害以降の処理(データの記

憶等)は、全て生存しているキャッシュを用いて行う。

【0029】WR処理時、転送データを2つのキャッシュ(329、330)に記憶させ、MP1(312)はCHA302へ終了報告を行う。DKU334へは、その後MP2(325)が上位装置とは非同期にデータの書き込みを行う。2つのキャッシュ(329、330)にデータを記憶させデータの2重化をさせている為、一方のキャッシュに障害が発生しても書き込みデータが失われることはない。

【0030】キャッシュ1(329)、キャッシュ2(330)にバッテリー331、332を用いてバッテリーバックアップとすることにより、システムを停止(電源OFF)させ、再度立ち上げた後でも停止直前のキャッシュ内のデータが保持される。よってシステム立ち上げ直後からキャッシュからの読み出しが可能となり、従来は必要であったDKUからの読み出しを必要とせずシステムスループットの向上を図ることができる。又、書き込みデータもキャッシュ内に保持される為、システム立ち上げ後DKUにデータを書き込むことができる。不慮の電源OFF後のシステム立ち上げ時も同様に、電源OFF直前のキャッシュ状態が保持されている為、直ちにキャッシュを用いた高速処理を行うことが可能となる。DKUへの反映が終了していない書き込みデータが存在してもキャッシュ内に保持されている為、電源ON後DKUへ書き込むことができる。

【0031】本実施例によれば、一つのキャッシュに障害が発生してもキャッシュ性能をゼロとすることなく処理が続き、又電源のOFF、ON直後からキャッシュ内に保持されたデータを用いて、キャッシュ性能を完全に引き出した処理を行うことができる。

【0032】また本実施例では、ディスクキャッシュを2分割としているが、3分割以上にしてもよく、N分割とすることにより、障害時の性能低下を1/Nとすることができる。

【0033】さらに本実施例では、ディスクキャッシュ全体を不揮発としているが、1部だけでもよい。従来装置同様に、WRデータのみを不揮発とする場合、通常、ディスク装置の運用状態が読み出し、書き込み比1/4

という現状を吟味すると、全体の1/5を不揮発とすることにより、データは保障される。

【0034】本実施例を適用することにより、ディスクキャッシュを搭載したディスク制御装置内の1つのディスクキャッシュで障害が発生しても、生存している他のディスクキャッシュにより大幅に性能を低下させることなく処理を続行できる。又、障害ディスクキャッシュの回復を行うことができ、従来技術と比べシステム性能を向上させることができる。システム停止、又は不慮の電源OFF後のシステム再立ち上げ直後から、ディスクキャッシュ内に残った電源OFF直前のデータを用いて、高速データ転送を行うことができる。尚、24時間運転システムへの適用は効果的である。

【0035】

【発明の効果】本発明におけるディスクキャッシュを搭載したディスク制御装置を含む記憶システム内において、ディスクキャッシュの機能を損なうことなく、継続して使用することができる。

【図面の簡単な説明】

【図1】実施例におけるシステム構成図

【図2】従来例のシステム構成図

【図3】従来例のシステム中に適用された、簡単化したトラックフォーマット

【符号の説明】

101、301…中央処理装置

102~109、302~309…チャンネル

110、310…ディスク制御装置

111、120、311、320…チャンネルスイッチ

128…バッテリーバックアップされて不揮発なメモリ(NVS)

112~119、124~127、312~319、3

25~328…マイクロプロセッサ

129、329、330…ディスクキャッシュ

331、332…バッテリー

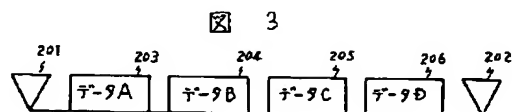
131、334…磁気ディスク装置

201、202…インデックスマーカ

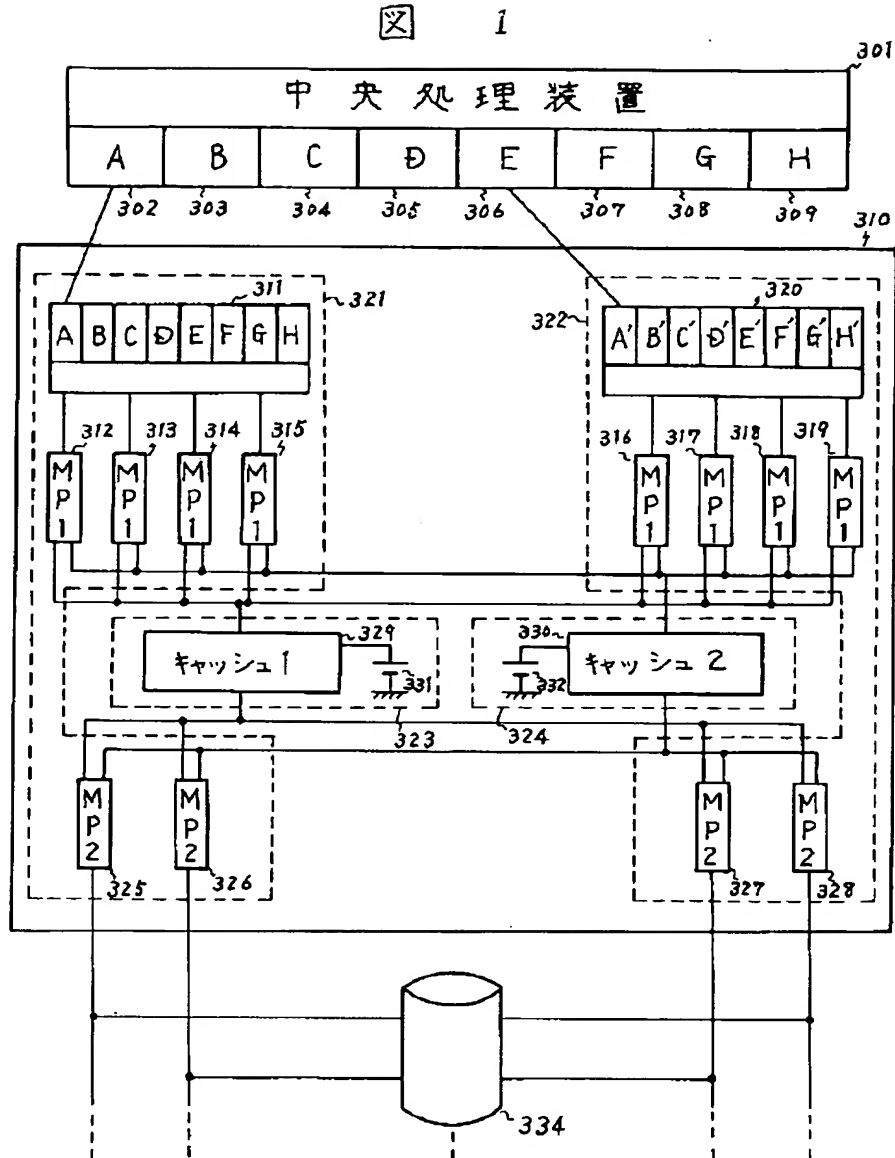
203~206…トラック上のデータ

121~123、321~324…電源領域

【図3】



【図1】



【図2】

フロントページの続き

(72)発明者 高松 久司
 神奈川県小田原市国府津2880番地株式会社
 日立製作所小田原工場内

(72)発明者 川口 幾雄
 神奈川県小田原市国府津2880番地株式会社
 日立製作所小田原工場内